

# パターン認識と機械学習

## 2.5 ノンパラメトリック法

05T4074A

久保田 敦

# 目次

- 目的
- ノンパラメトリック法
- ヒストグラム密度推定法
- まとめ

# 目的

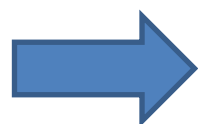
- パラメトリックとノンパラメトリックそれぞれの特徴を改めて理解し、またノンパラメトリックなアプローチというものを理解する。
- ヒストグラム密度推定法について学ぶ。

# ノンパラメトリック法

パラメトリック: 少数のパラメータから確率変数の分布の形状を決まる

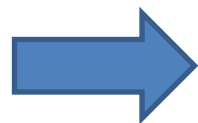
ノンパラメトリック: 分布の形状が制限されず、データによって形状が決まる

パラメトリックなアプローチ  
分布の形状を仮定



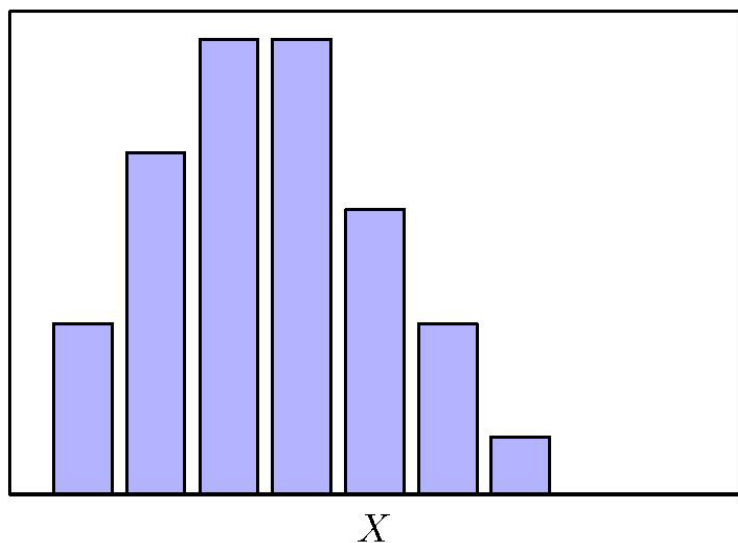
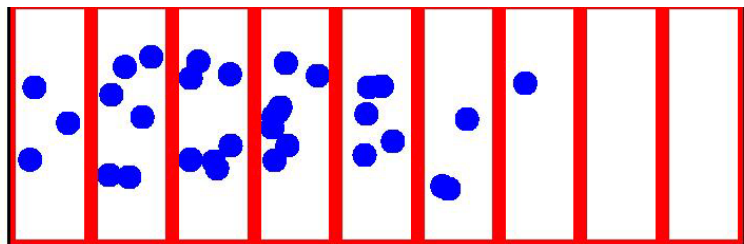
制限がかかり  
予測性能が悪くなりうる

ノンパラメトリックなアプローチ  
確率密度関数の形が  
データに依存して決まる



制限がない

# ヒストグラム密度推定法



$x$  : 連続変数

$\Delta_i$  :  $i$ 番目の幅

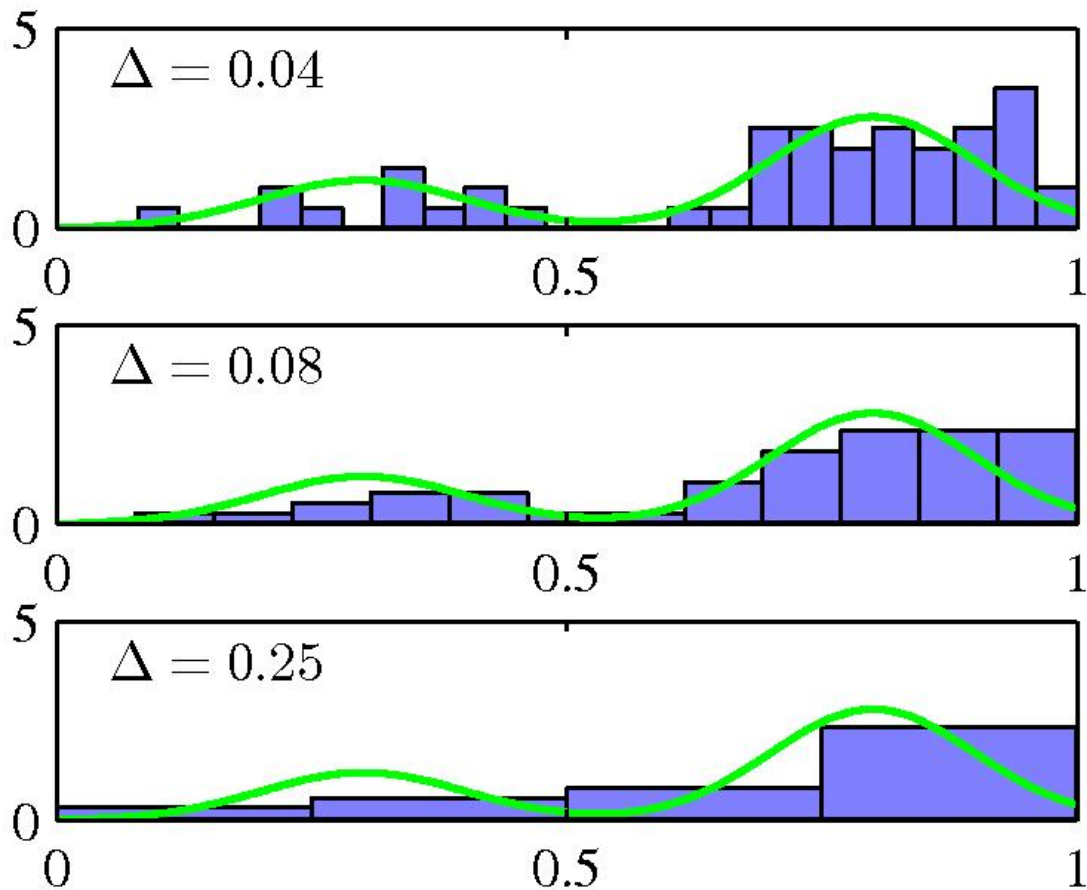
$n_i$  :  $i$ 番目の観測値の数

$N$  : 観測値の総数

$$\text{確率密度} : p_i = \frac{n_i}{N\Delta_i}$$

# $\Delta$ の与える影響

$\Delta$ の値は適当な値にしないと分布の特徴をとらえきれない。



# ヒストグラム密度推定法(補足)

利点	欠点
<ol style="list-style-type: none"><li>一度ヒストグラムを求めると元データを破棄できる。 (大規模なデータに有利)</li><li>データが逐次的に与えられた時に容易に適用できる。</li></ol>	<ol style="list-style-type: none"><li>推定した密度が区間の縁で不連続になる。</li><li>次元数が増えると、指数的に区間の総数が増え、計算規模が増加する。</li></ol>

実際には、1次元か2次元のデータの簡単な可視化には役立つが、他のほとんどの密度推定の応用問題には適してない。

# まとめ

## ノンパラメトリックなアプローチでは

1. 特定の位置の確率密度を推定するにはその点の近傍のデータ点も考慮する。
  - 近傍の特性は区間によって定義されている。  
(区間 = ヒストグラムアプローチの平滑化パラメータ)
2. 平滑化パラメータの値は大きすぎも小さすぎもしない、適切な値でないといけない。